Available online at www.sciencedirect.com
SCIENCE DIRECT
European Journal of Medicinal Chemistry 41 (2006) 862–873
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://france.elsevier.com/direct/ejmech

Original article

QSAR models based on quantum topological molecular similarity

P.L.A. Popelier *, P.J. Smith

School of Chemistry, Sackville Site, North Campus, University of Manchester,

Manchester M60 1QD, UK

Received in revised form 1 March 2006; accepted 20 March 2006
Available online 12 May 2006

Abstract

A new method called quantum topological molecular similarity (QTMS) was fairly recently proposed [J. Chem. Inf. Comp. Sc., 41, 2001, 764] to construct a variety of medicinal, ecological and physical organic QSAR/QSPRs. QTMS method uses quantum chemical topology (QCT) to define electronic descriptors drawn from modern ab initio wave functions of geometry-optimised molecules. It was shown that the current abundance of computing power can be utilised to inject realistic descriptors into QSAR/QSPRs. In this article we study seven datasets of medicinal interest [1]: the dissociation constants (pK_a) for a set of substituted imidazolines [2], the pK_a of imidazoles [3], the ability of a set of indole derivatives to displace [^{3}H] flunitrazepam from binding to bovine cortical membranes [4], the influenza inhibition constants for a set of benzimidazoles [5], the interaction constants for a set of amides and the enzyme liver alcohol dehydrogenase [6], the natriuretic activity of sulphonamide carbonic anhydrase inhibitors and [7] the toxicity of a series of benzyl alcohols. A partial least square analysis in conjunction with a genetic algorithm delivered excellent models. They are also able to highlight the active site, of the ligand or the molecule whose structure determines the activity. The advantages and limitations of QTMS are discussed.
© 2006 Elsevier SAS. All rights reserved.

Keywords: Ab initio; PLS; p*K*a; QSAR; QSPR; Electron density; Topology; Atoms in molecules; Quantum chemical topology

1. Introduction

Many biochemical and physicochemical properties are successfully predicted by descriptors drawn from quantum chemistry [1,2]. A number of recent examples support this assertion. For instance, TAE (transferable atom equivalent) [3,4] electron-density-based descriptors were used for the prediction [5] of protein retention times in anion-exchange chromatography systems and the a priori prediction of protein adsorption behaviour [6], molecular orbital indices enabled the forecast of mutagenic activity of furanones [7], atomic charges featured as descriptors in the prediction of $\log(1/IC_{50})$ values of nifedipine analogues [8], EEM [9] -derived descriptors generated structure–activity relationships for the prediction of metabolism by human UDP-glucuronosyltransferase isoforms [10], minimum molecular surface local ionisation energies correlated to the pK_a of substituted anilines [11], ab initio wave functions were used in the prediction of intestinal drug permeability [12], Hartree–Fock derived dipole moments played a part in the classification of aroma compounds [13], electrostatic potentials contributed to a docking study of the CCR2 receptor [14], frontier molecular orbitals of flavones aided in the rationalisation of a plausible binding to GABA receptors [15], structure–property relationships between Hammett σ constants and momentum space quantities were set up [16], and molecular quantum similarity measures were successfully used in numerous studies [17,18], with extension to fragment-based quantum similarity measures [19]. This handful of examples already gives an idea of the wide range of activities and properties profiting from a considerable variety of quantum chemical descriptors.

One straightforward quantum mechanical entity that seems to have been overlooked as a descriptor in its own right is the geometry-optimised bond length. Of course, several research groups take geometry-optimised molecules (local minima in the potential energy surface) as a starting point for subsequent analysis. However, the actual bond lengths themselves do not feature as descriptors. For the last decade computers have been

* Corresponding author. Tel.: +44 161 306 4511; fax: +44 161 306 4559.
E-mail address: pla@manchester.ac.uk (P.L.A. Popelier).

0223-5234/$ - see front matter © 2006 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.03.004

powerful enough to optimise molecules up to a molar mass of 500 g mol^{-1} by semi-empirical methods such as AM1. More recently, such molecules can be routinely optimised at (still modest but reasonable) ab initio levels of theory (such as HF/3-21G and HF/6-31G*). Recently, work [20] carried out in our group showed that optimised bond lengths are effective in predicting the corticosteroid-binding globulin (CBG) binding activity of the classic steroid data set of Cramer et al. [21], and in predicting the antibacterial activity of nitrofuran derivatives. Using partial least squares (PLS) in conjunction with a genetic algorithm (GA) it proved fruitful to set up a QSAR that applied (optimised) bond lengths as descriptors. Although a simple finding, the indirect path that led to it involved a theory adequately called [22] quantum chemical topology (QCT) [23–26]. Other than being a generalisation of quantum mechanics to subspaces [27–29] QCT is widely appreciated as a theory that extracts chemical insight from sufficiently accurate wave functions, as can be seen from two recent reports on hundreds of papers by research groups using QCT [30,31].

As explained in more detail in the Section 2, QCT defines special points in the total molecular electron density, characterising bonds. As such, QCT offers a well-defined mathematical procedure to represent a molecule by a discrete "quantum fingerprint", consisting of a string of numbers. Because of their compactness, these fingerprints allow a rapid comparison between molecules, and hence form the basis for measures of similarity. The notion of QCT descriptors (details in Section 2) was first proposed (32) in 1999, in the context of QSAR/QSPR. This work showed that it was not necessary to compare complete electron densities of benzoic acids in order to predict their acidity. Instead, quantum mechanical properties evaluated at specific points in space, as prescribed by QCT, sufficed.

Soon after, the QCT descriptors were used in a supervised context, and delivered a set of QSARs of medicinal [20,33,34] and ecological [35–38] interest and QSPRs [36,39–42]. Evidence has been accumulated that QCT descriptors are selectively capturing electronic effects, which renders them in a good position to replace Hammett constants [42,43] (including σ^+ and σ^-). An important advantage is that QCT descriptors are always available even for functional groups for which Hammett constants are lacking. In fact, according to an extensive study [44], only 63 of the 100 common substituents taken from the *logpstar* database had measured Hammett sigma constants.

Surely, QCT descriptors can be combined with log *P* and steric parameters, which constitute the two remaining types of descriptors. Although we do not expand on this here, one can imagine applications where it is relevant to know if the measured activity is actually electronic in nature. QCT descriptors can unambiguously decide on this matter since they do not give rise to a good QSAR model if either log *P* or steric effects alone are governing the activity in question.

The aim of this article is to demonstrate the application range (or "action radius") of QCT descriptors. For many of the seven investigated data sets improved regression statistics emerged compared to other methods such as CoMFA, electrotopological indices [45,46] and quantum molecular similarity indices.

An interesting feature of the current method is that it is able to localise a part in the molecule where the chemical change associated with the observed activity actually happens. For example, the O–H bond is highlighted [36] as the "active site" in the deprotonation of carboxylic acids if their acidity is taken as the observed data. Obviously here we recover what is expected but in the case of the antitumour activity of phenylbutenones the active site analysis confirmed [33] an earlier hypothesis that these compounds act via a Michael addition. In another case [34], in the context of mutagenic activity, this type of analysis (details see next section) suggested a preferred mechanistic pathway for the initial hydroxylation of dimethyl heteroaromatic triazenes, an issue that had remained ambiguous. In this work, we continue to explore the use the active site analysis and contrast it with literature results.

2. Methods and materials

2.1. Data sets

All activity data were sampled from the literature and repeated for convenience in the Annex A. Table 1 summarises the seven datasets subject to the current method. The smallest number of compounds is 14 and the largest 23, totalling 126 molecules over all sets. The sets have been selected on the grounds that electronic descriptors feature strongly (over steric parameters and log *P*), and that we can compare with the results of alternative methods. We have performed analyses on several sets of compounds where the mechanism is relatively well known, or at least the factors considered important for the mode of action have been delineated.

2.2. Quantum chemical topology

Molecular electron densities contain special points where the gradient of the electron density, denoted by $\nabla\rho$, vanishes. They are called critical points. There are four types of them, depending on the nature of the curvature of ρ in their vicinity. One type of critical point is of relevance for this work. It is called a bond critical point (BCP) since it appears in between

Table 1
Survey of the analysed data sets

n°	Data set	*n* [a]	Activity	Reference
1	Imidazolines	23	Dissociation (pK_a)	[65]
2	Imidazoles	15	Dissociation (pK_a)	[69]
3	Indoles	23	Displace [^{3}H] flunitrazepam from binding to cortical membranes	[70]
4	Benzimidazoles	15	Influenza inhibition constants	[76]
5	Amides	17	Inhibition of liver alcohol dehydrogenase	[77]
6	Sulphonamides	19	Inhibition of carbonic anhydrase	[78]
7	Chlorophenols	14	Toxicity to submitochondrial particles	[79]

[a] Number of molecules in the set.

two nuclei that are bonded (for a more precise description and discussion see [47]). BCPs can be efficiently located by a special algorithm [48,49]. Certain functions evaluated at a given BCP characterise the corresponding bond. We have used four such functions, namely, the electron density, the ellipticity, the Laplacian of the electron density, and the kinetic energy density K(**r**). Before describing them in turn, the curvature of the electron density around a BCP needs to be clarified. The Hessian matrix contains this information. It is a symmetric three-by-three matrix containing second derivatives as elements, that is $\partial^2\rho/\partial q_i \partial q_j$, where q can be *x, y* or *z*. Two eigenvalues of the Hessian, denoted by λ_1 and λ_2, are negative and measure the curvature perpendicular to the bond. The positive eigenvalue, λ_3, measures the curvature along the bond.

The electron density at the BCP, ρ_b, can be related to bond order via an exponential relationship [50]. Although initially proposed for carbon–carbon bonds this interpretation was extended [51] for a wide variety of bonds via a linear relationship of the type $a + b\rho_b + c\lambda_3 + d(\lambda_1 + \lambda_2)$, where *a, b, c* and *d* are fitted parameters. In that work [51], ρ_b and λ_3 were interpreted as measures of σ character, whilst $(\lambda_1 + \lambda_2)$ measured the degree of π character. For the latter we employ an alternative measure here, called the ellipticity and defined as $\varepsilon = (\lambda_1/\lambda_2) - 1$. In a more general context the ellipticity is a measure of structural stability [52] but it has traditionally been used as a measure of π character in homopolar bonds. In spite of its variation along polar bonds [53] it can safely be used when evaluated at the BCP because of the small changes in its position for a given bond going from one substituted compound to another. The third function is the Laplacian, $\nabla^2\rho$, which can be defined as $\lambda_1 + \lambda_2 + \lambda_3$. If negative, it indicates that electronic charge is locally concentrated. Often it features as a simple measure of covalent character. Finally, the fourth function is K(**r**), is defined as

$$K(r) = -\frac{1}{4}N\int d\tau' \ [\psi^* \ \nabla^2\psi + \psi \ \nabla^2\psi^*] = -\frac{1}{2}N\int d\tau' \ \psi \ \nabla^2\psi \quad (1)$$

where ψ is the many-electron wave function (which is always real in this work) and $\int d\tau'$ denotes an integration over the spin coordinates of all N electrons except one. Interpreting K(**r**) in chemical terms is not straightforward although useful formulae describing its link to the Laplacian and the more "classical" kinetic energy G(**r**) can be found elsewhere [54].

2.3. Computational details (ab initio calculations and statistics)

Our method, which has been coined quantum topological molecular similarity (QTMS) essentially proposes a chemometric analysis using QCT descriptors [55]. Many details of an earlier version of this method have been published before [56].

All molecules were geometry-optimised and their wave functions computed at the HF/6-31G*//HF/6-31G* level of theory [57] using the GAUSSIAN98 suite of programs [58]. Earlier work on other systems, carried out in our lab, made us conclude that this level of theory is a good compromise between speed and reliability. However, due to their size the indole derivatives were optimised at a lower level, HF/3-21G*//HF/3-21G*, which we deem the lowest pure ab initio level of theory reliable enough for our purpose. The work was carried out on a very small PC cluster (10 AMD Athlon 1.8 GHz/1Gb RAM nodes).

Subsequently the QCT program MORPHY98 [59] located the BCPs in molecules and evaluated, at each BCP position, the four functions described in the previous subsection. The computation of the QCT descriptors only *takes a very small fraction* of the CPU time required to obtain the optimised wave functions. The computation of BCP properties should not be confused with that of QCT's atomic properties, which is indeed expensive.

In the next stage, a PLS analysis [60], carried out by the program SIMCA-P(61), fits the QCT descriptors to the activities. We adopted SIMCA-P's predefined criterion for determining the significance of a latent variable (LV). If the q^2 of a newly constructed LV is less than 0.097, then that variable is unimportant to the model and is discarded. No more LVs are computed once this occurs and the PLS regression is completed. The PLS analysis yields four statistics to judge the quality and validity of the model. The first two statistics provide a measure of the quality of the model and are the correlation coefficient, r^2, and the cross-validated correlation coefficient, q^2. The final two statistics obtained provide a safeguard against the possibility that the model may have been obtained by chance. This is referred to as a randomisation test. The activities (i.e. dependent variables) are repeatedly randomised, at least 10 times, and a fresh PLS analysis is carried out for each such data permutation. The dependent variable is randomly reordered whilst the descriptor variables are left alone. Hence, each set of descriptors is deliberately associated with the wrong activity, with the intention of finding out if the descriptors fit the activity "for the right reasons". The subsequent PLS analysis will produce new values of r^2 and q^2 and if these values are lower than those of the original model it is likely that the relationship was not obtained by chance. Each new r^2 and q^2 value (one for each permutation) is plotted against the absolute value of the correlation coefficient between the original response variable and its permutation. A line is then fitted through the r^2 values and another separate line through the q^2 values. One then examines the intercept of each line, denoted by r^2(int) and q^2(int). For a model to be valid it is recommended [61] that r^2(int) be lower than 0.4 and q^2(int) lower than 0.05. Note that it is perfectly possible that r^2(int) and q^2(int) are negative.

In order to obtain a valid model it is sometimes necessary to gradually reduce the number of LVs until the randomisation test is passed. This is justified as there is no point in attempting to interpret a model, even it does possess high values of r^2 and q^2, if it has been obtained purely by chance.

In addition, we used a genetic algorithm to select the BCP properties to include in the regression in order to optimise the regression. For variable selection, the MATLAB genalg.m routine from the PLS_toolbox [62] was employed, using a population size of 256, a mutation rate of 0.003, and a maximum number of generations of 200. The fitness function was taken as the q^2 value from a PLS analysis performed on the dataset using variables selected by the GA. The following abbreviations will be used for labelling the BCP properties selected by the GA: rho (the value of the electron density), lap (the value of the Laplacian of the density), ell (the value of the ellipticity), and K (the value of the kinetic energy density). Each property will be followed by the numerical labels of the atoms which constitute the bond, e.g. if the GA selects variable lap2_3, this means the value of the Laplacian at the BCP between atoms 2 and 3 (according to the provided numbering schemes) has been chosen.

The restrictions on the datasets are the same as for typical Hansch type QSARs, namely the dataset must be congeneric (i.e. a common structural skeleton differing in only the substituents must be present), the mode of action for all the molecules must be the same, and the dataset should be large enough to avoid possible chance correlations being obtained between the descriptors and activities.

2.4. Active site analysis

The program SIMCA offers a module that computes the Variables Importance in the Projection (VIP) [60]. The VIP values reflect the importance of terms in the PLS model with respect to *Y*, i.e. its correlation to all responses, and with respect to *X*. Variables with higher VIP scores are more relevant in explaining the activity. VIP values are typically plotted as histograms, which are shown in many Figures in this paper. The *X* variables (here GA-selected QCT descriptors) with the highest VIP values are associated with the active part of the molecule. Ideally, the profile of the VIP histogram displays a sharp decline after the first few variables with the highest VIP values. In that case, the active site is well localised provided the QCT descriptors are associated with a cluster of *neighbouring* bonds. If the VIP profile drops off slowly then the active site is poorly localised or diffuse. Note that the term "active site" is commonly understood to refer to the binding cavity in an enzyme but here it refers a highlighted zone in the *ligand* or the molecule whose structure determines the activity.

Another problem is that this method can suffer from "contaminations" since sometimes variables with high VIP scores cannot be related to the activity. For example, in an active site analysis of a sizeable set of carboxylic acids [36] the highest VIP value corresponds to the O–H bond. This is expected since this bond breaks and reforms in the establishment of the equilibrium of the acids with their respective ionic forms. The second VIP value corresponds to the C-O bond in the carboxyl group, which is again understandable given its proximity to the OH bond. Curiously, the third VIP value is the *meta* CH bond. It is difficult to reconcile this bond with the mechanism for acid dissociation. We note that in their study on the identification of active molecular sites using quantum-self-similarity measures Amat et al. [63] also detected unexpected fragments correlating strongly with σ constants. These authors suggested a possible explanation for this phenomenon based on Mezey's "holographic electron density theorem" [64].

3. Results and discussion

3.1. Dissociation of imidazolines

Table S1 of Annex A contains a list of substituted imidazolines reported in a paper by Good et al. [65] and ultimately adopted from the experimental work by Timmermans and Zwieten [66]. The activity data refer to the pK_a of the compounds. Fig. 1 shows a representation of the common molecular skeleton. This class of molecules received much attention as they are similar to clonidine, which is a very active hypertensive drug. It was shown [67] that the dissociation constant of clonidine analogues plays an important role for the activity. Included in Table 2 are the results of the PLS analysis and the observed versus predicted plot is given in Fig. 2. Electronic effects are of crucial importance in determining the activity of this dataset.

Using electrostatic potential and electric field similarity calculations, Good et al. obtained an optimal cross-validated r^2 value of 0.96. Excellent statistics were also found by Kim and Martin [68] who performed a CoMFA analysis on the set, producing an r^2 value of 0.98 (the authors did not provide a q^2). It is gratifying to note that QTMS produces improved statistics compared to both these methods (although all the models are outstanding), and adds weight to the conjecture that our approach favours datasets where electronic effects dominate. The VIP plot for the BCP properties is given in Fig. 3.

In this set of compounds, QTMS highlights the exact region one would expect to be important in determining the pK_a values. For the imidazoline molecules, one would expect protonation to occur on the basic nitrogen, atom 6 in Fig. 1, which is the same region given priority in the VIP plot. There is a noticeable drop in the VIP scores after the first 12 variables, and out of these variables, five are associated with the two bonds involving N_6, namely C_7–N_6 and N_6=C_2. More precisely, five out of the first six variables with the highest VIP scores are associated with the basic nitrogen N_6. For this set of compounds, QTMS is able to produce excellent and validated re-

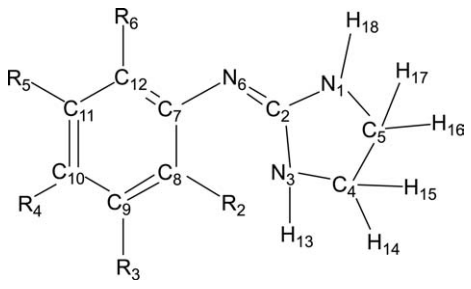

Fig. 1. Numbered common skeleton for the set of imidazolines.

Table 2
Regression statistics for the PLS analysis of all data sets

n°	Molecules	LVs	r^2	q^2	r^2(int)	q^2(int)	Comment
1	Imidazolines	2	0.995	0.993	0.10	–0.18	
2	Imidazoles	3	0.989	0.956	0.14	–0.25	
3	Indoles	3	0.821	0.623	0.09	–0.15	
4	Benzimidazoles	4	0.913	0.797	0.30	–0.13	
5	Amides	2	0.761	0.596	0.06	–0.14	log *P* excluded
		3	0.923	0.734	0.17	–0.16	log *P* included
6	Sulphonamides	2	0.910	0.891	0.07	–0.10	
7	Chlorophenols	2	0.928	0.840	0.20	–0.16	

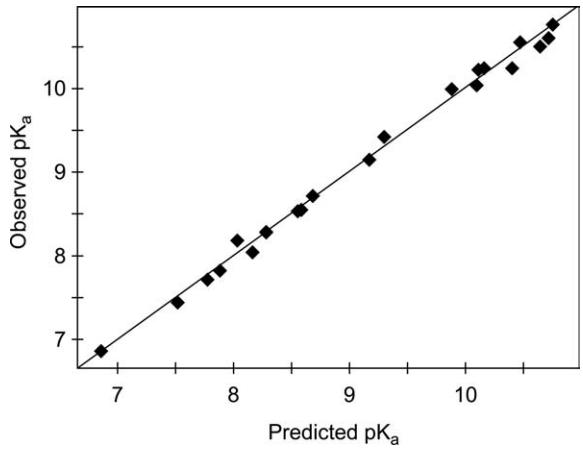

Fig. 2. Observed versus predicted pK_a values for the set of imidazolines.

gressions, and highlights the portion of the molecule responsible for the observed activity.

3.2. Dissociation of imidazoles

Imidazole can be protonated due to the presence of a pair of non-bonding electrons that are not involved in resonance. The conjugate acid of imidazole has a pK_a of 6.8 and exists in both the protonated and unprotonated forms at physiological pH. This is one of the reasons that histidine, an imidazole-containing amino acid, is an important catalyst in many biological reactions. The cumbersome and time-consuming ways of obtaining pK_a values experimentally mean that the ability to reliably predict pK_a values is of considerable importance. Many different approaches have been applied to the problem, most notably CoMFA, which has achieved considerable success in these kinds of applications [68]. Here we apply QTMS to a set of 15 imidazole derivatives, the substituents of which are given in Table S2. This set has been examined by Silverman and Platt [69] in a comparative molecular moment analysis (CoMMA) study. This method attempted to correlate pK_a with similarity descriptors based on moments of the mass and charge distribution of the molecules. Fig. 4 gives the common skeleton of the molecules in the dataset. The results of the PLS analysis are displayed in Table 2. A plot of observed versus calculated pK_a values is given in Fig. 5.

It is clear from the regression statistics obtained that BCP properties are capable of producing a model with an outstanding goodness of fit and excellent predictive capabilities. The results obtained are comparable with a CoMFA study performed by Kim and Martin [68], who obtained a correlation coefficient of 0.99 (the authors did not provide a cross-validated statistic), and improved compared to the results obtained by Silverman and Platt, who obtained a maximum predictive cross-validated r^2 value of 0.772. This set has also been examined by Good et al. [65], who used PLS analysis with steric and electrostatic similarity matrices for the compounds, yielding a maximum q^2 value of 0.91.

Turning our attention to the active site of the molecules, the VIP values are given in Fig. 6. One would expect the area surrounding atom N_2 to be highlighted as being important since this is the atom that is involved in protonation. However,

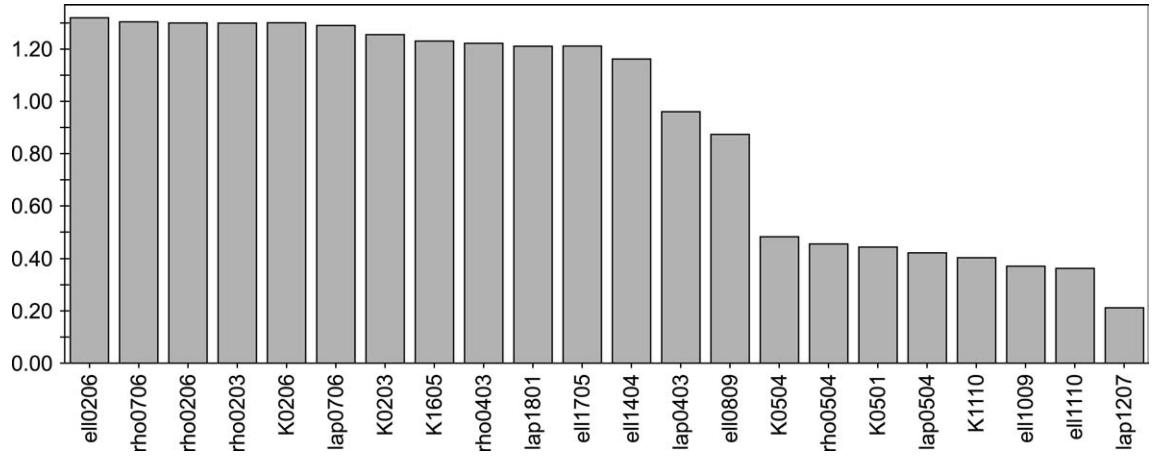

Fig. 3. VIP plot for the variables selected in the PLS analysis of the set of imidazolines.

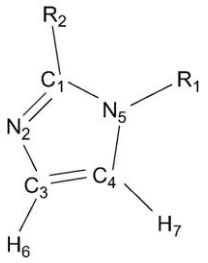

Fig. 4. Numbered common skeleton for the set of imidazoles.

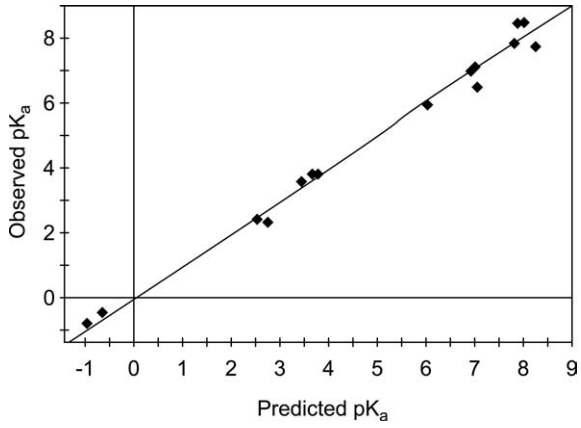

Fig. 5. Observed versus predicted pK_a values for the set of imidazoles.

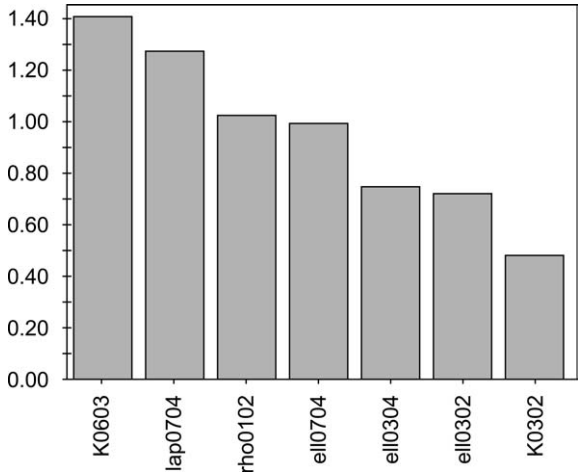

Fig. 6. VIP plot for the variables selected in the PLS analysis of the set of imidazoles.

bond properties involving this atom take only third, sixth and seventh place in the histogram of Fig. 6. Given that only the third one is significant, it is appropriate to admit that the VIP plot is not able to highlight the place where the reaction site occurs. Perhaps a series of protonated moieties need to be analysed, introducing a N_2–H bond.

3.3. Inverse agonist activity of indole derivatives

In this example we investigate a set that was analysed [70] earlier by an alternative quantum similarity measure. The set

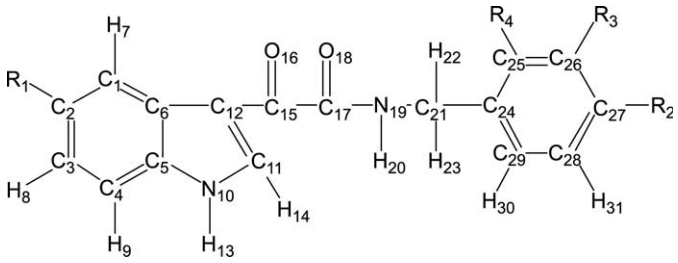

Fig. 7. Numbered common skeleton for the set of indoles.

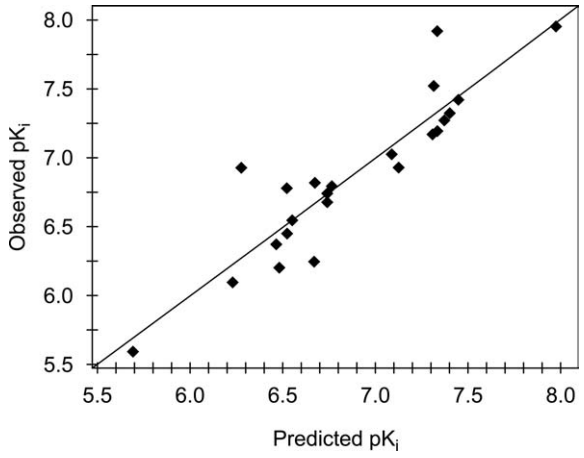

Fig. 8. Observed versus predicted pK_i values for the set of indoles.

consists of 23 indole derivatives, able to displace [^{3}H] flunitrazepam from binding to bovine brain membrane. These derivatives are known to bind to the benzodiazepine receptor and due to the wide-spread employment of benzodiazepines in clinical practice (sedatives, anxiety treatment and anticonvulsants) the compounds are of substantial pharmacological interest [71]. Fig. 7 gives the common skeleton of the indole structure and Table S3 contains the structure and activity data taken from Ref. [72]. Here the activity is measured by pK_i, where K_i is the concentration necessary to produce 50% inhibition in an assay with bovine brain membrane. Table 2 shows the regression results and Fig. 8 the observed versus predicted plot for the activity.

This set of compounds has already been examined by Hadjipavlou-Litina and Hansch using traditional descriptors [72]. Their main conclusion was that hydrophobic effects are unimportant for this dataset and that the decisive role in determining activity is electronic in nature. The best model the authors obtained consisted of the Hammett sigma parameter and two indicator variables, yielding an r^2 value of 0.81 (no value of q^2 was provided). It should be noted that three compounds (2, 13 and 14) were excluded from this regression. In contrast, QTMS produces a slightly better and validated r^2 value (Table 2) statistic *for the full set* of derivatives. In another study [70] the Carbó index was employed to model the activity of the indole molecules. The researchers obtained comparable statistics to ours, with an $r^2 = 0.761$ and a $q^2 = 0.631$, but again, three compounds (1, 13 and 14) were excluded. When the same compounds were excluded in a repeat QTMS analysis a model

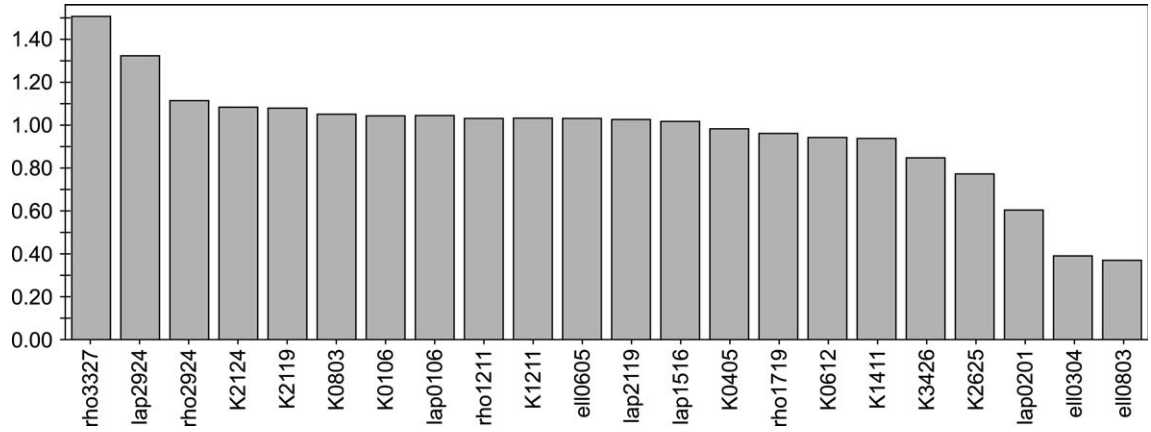

Fig. 9. VIP plot for the variables selected in the PLS analysis of the set of indoles.

was obtained with $r^2 = 0.855$ and $q^2 = 0.743$. In a follow-up to this study [73], electron-electron repulsion energies were employed by the same group to obtain a model with $r^2 = 0.603$ and $q^2 = 0.507$, now including all 23 compounds. It is clear that QTMS yields better predictions.

The VIP plot from the original QTMS study (featuring all 23 molecules) is given in Fig. 9. A pharmacophore model by Cook and co-workers (Zhang et al. [74]) consists of the following interaction sites: (i) a hydrogen bond acceptor site (indole's $N_{10}H_{13}$ group) (ii) a hydrogen bond donor site (C_{17}=O_{18}), (iii) a "bifunctional" hydrogen bond donor/acceptor site (C_{15}=O_{16}), and (iv) three lipophilic pockets (C_{21} methylene and near aryl ring). In the VIP plot only two bonds stick out, $C_{27}X_{33}$ (where X = H, O or Cl; see R_2 substituents in Table S3) and $C_{24}C_{29}$, both part of the aryl ring. It is tempting to link this highlighting region with a hypothesis forwarded by Da Settimo et al. [71], which is that recognition of the aryl moiety by a receptor subsite, "drives" the anchoring of the rest of the ligand structure.

3.4. Influenza virus inhibition of benzimidazoles

Tamm et al. [75] have reported the influenza virus inhibitory activity for a series of alkylated benzimidazoles, the common structure of which is shown in Fig. 10. Table S4 lists the substituents and the activity data, represented by the inhibition constant, pK_i. Table 2 contains the results of the PLS analysis and Fig. 11 shows the observed versus predicted plot for the activities.

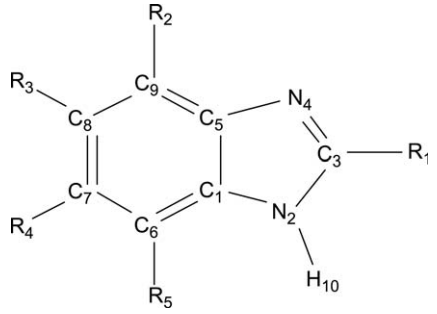

Fig. 10. Numbered common skeleton for the set of benzimidazoles.

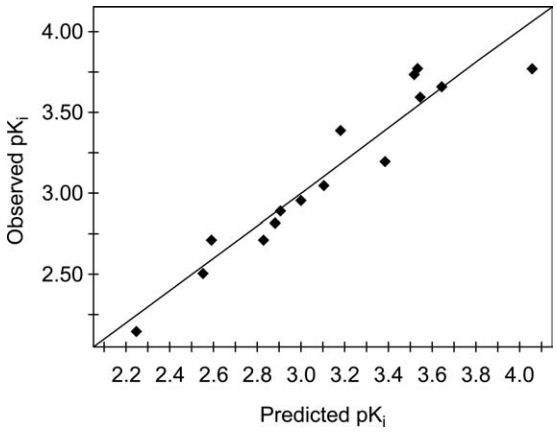

Fig. 11. Observed versus predicted pK_i values for the set of benzimidazoles.

A satisfactory model is obtained for the dataset, although the value of the r^2 (int) statistic appears quite high. This is probably due to the low ratio of compounds to latent variables, although the model itself does pass the randomisation test and can be accepted as valid. The VIP plot for the selected BCP properties is given in Fig. 12.

This set of molecules has been studied by Hall et al. [76] who modelled the inhibition activity using E-State indices. Using the E-State value of C_3 (see Fig. 10) and the averaged E-State values of N_2 and N_4 as molecular descriptors, a model with $r^2 = 0.91$ was obtained. The authors conclude that substitution on C_3 is more important in determining the activity than any other position, and that the five-membered ring plays a significant role at the receptor. In terms of the region picked out by QTMS, our results are in agreement with the E-State investigation. Indeed, eight out of the 11 selected variables (Fig. 12) are BCP properties associated with bonds in the five-membered ring, with the top four VIP scores belonging to BCP properties associated with this region of the molecule.

3.5. Inhibition of alcohol dehydrogenase by amides

Alcohol dehydrogenase (ADH) is a zinc-containing enzyme that can oxidise a variety of alcohols to aldehydes or reduce

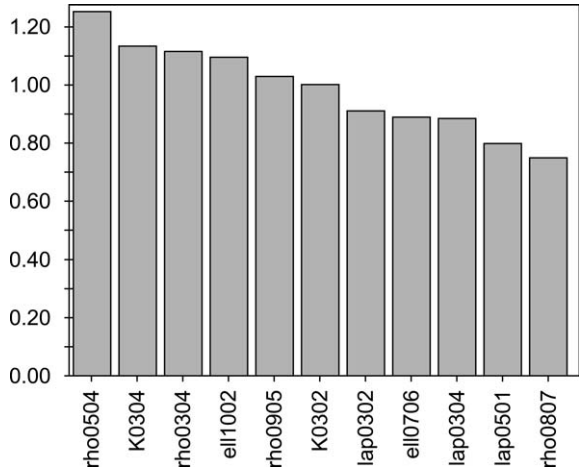

Fig. 12. VIP plot for the variables selected in the PLS analysis of the set of benzimidazoles.

aldehydes to alcohols, depending on the reaction conditions. The mechanism of the enzyme has considerable medicinal interest as a possible antidote to methanol poisoning, and has also been studied to better understand the problem of alcoholism. In an attempt to delineate the hydrophobic, electronic, and steric roles of substituents in the interaction of the enzyme with small molecules, a number of QSAR have been derived by Hansch et al. [77] for the inhibition of ADH. In an effort to investigate the suitability of QTMS in accounting for electronic effects, a set of amide ADH inhibitors was studied. The structural and interaction constants are presented in Table S5 and the common skeleton is given in Fig. 13.

In order to assess the ability of QTMS to act as an electronic descriptor, and to compare our results with those of Hansch, two QTMS analyses were performed: one where the log *P* values of the compounds were permitted to be selected by the GA as descriptor variables, and a second analysis where log *P* was excluded. The results of both PLS regressions are given in Table 2, and the observed versus predicted plot for the analysis with log *P* included is given in Fig. 14.

It is interesting to note that the GA selects log P when this variable is included, and that inclusion of this parameter leads to a significant improvement in the regression statistics. Hence lipophilic effects play a part in the inhibition of the enzyme. This conclusion is in agreement with that of Hansch, who found electronic and lipophilic effects to be dominant in this dataset, with negligible contributions from steric effects. Using log P and the Hammett σ* constant, Hansch et al. produced an equation with $r^2 = 0.738$. This equation was only for 15 compounds (owing to poor experimental data for compound **10** and

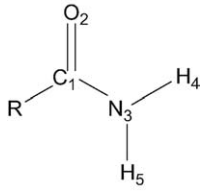

Fig. 13. Numbered common skeleton for the set of amides.

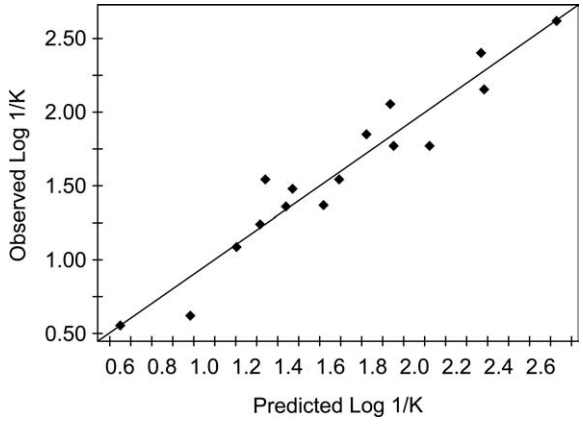

Fig. 14. Observed versus predicted log(1/K) values for the set of amides.

11), however, and when a repeat QTMS analysis was performed for this number (using the same selected variables), a validated model was obtained with $r^2 = 0.939$, and $q^2 = 0.872$. These results are better compared to those obtained using the classical parameters, and demonstrate that, in this case, not only do BCP properties adequately model the electronic part of the QSAR, but actually improve the results. It can also be seen that the BCP properties can satisfactorily be used in conjunction with other conventional descriptors.

The VIP plot for the variables is shown in Fig. 15. Hansch et al. concluded from the negative sign of the σ* coefficient in their regression equations that electron-releasing substituents favour the inhibition. This information, combined with the negligible steric effects, led Hansch and co-workers to the conclusion that it is the carbonyl group that is critical in binding the molecule to the enzyme. The most important bond according to the VIP plot, however, is the C-N bond adjacent to the carbonyl group, although the Laplacian of the carbonyl bond is also selected by the GA. The reason for this discrepancy is not clear although our results remain qualitatively the same if the GA is de-activated. Future work may involve a higher level of calculation, including electron correlation.

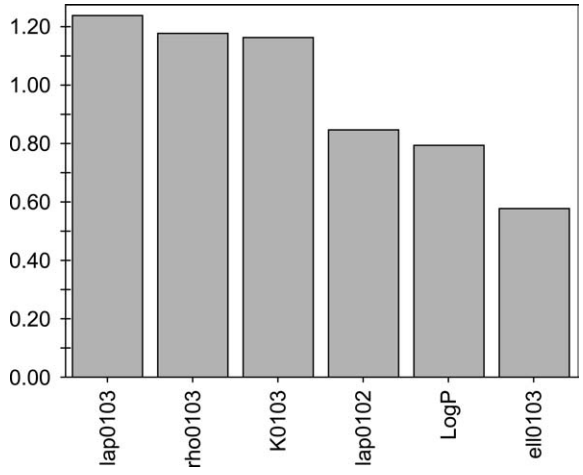

Fig. 15. VIP plot for the variables selected in the PLS analysis of the set of amides.

3.6. Inhibition of carbonic anhydrase by sulphonamides

Carbonic anhydrase (CA) is a zinc-containing enzyme that catalyses the hydration of CO_2 to H_2CO_3. Inhibitors of CA, particularly sulphonamide inhibitors, have attracted much interest from medicinal chemists, mainly for their diuretic properties and ability to treat glaucoma. The set of sulphonamide derivatives presented below have been studied via Hansch analysis by Kakeya et al. [78], using classical hydrophobic and electronic parameters. Owing to two very different common structures the set has been divided into two sets given in Fig. 16a, b. The corresponding activities are given in Tables S6a and S6b, respectively. The K_i symbol represents the CA inhibition constant. It should be noted that compounds 17, 18, and 19 (which are *ortho*-substituted) have been omitted from our analysis in order compare our statistics with those of Kakeya et al., who also excluded the compounds from analysis.

In order to model both sets simultaneously, the SO_2NH_2 moiety common to both sets of molecules has been considered separately and only BCP properties associated with bonds in this fragment are included in the QTMS analysis. The common skeleton for this group is given in Fig. 17. Table 2 contains the results of the PLS analysis and Fig. 18 shows the observed versus predicted plot for the activities.

Using a combination of classical descriptors, Kakeya et al. obtained regression models with r^2 values of 0.887 and 0.84, respectively. These values were obtained using the lipophilicity parameter, π, in combination with classical electronic parameters. The authors stated that combining the electronic parameters with lipophilic parameters produces more significant correlation than either of these parameters separately, and so despite our high correlations (Table 2) hydrophobicity is important for this dataset. Addition of the π parameter to our PLS analysis produces an improved model with $r^2 = 0.943$, and $q^2 = 0.914$. Since the addition of the hydrophobic effect improves our statistics (Table 2) only marginally we infer that the dominant factor is electronic in nature. This dataset again

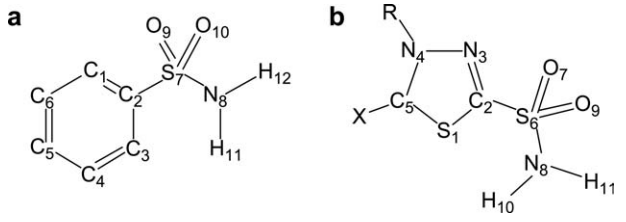

Fig. 16. Numbered common skeleton for the (a) first and (b) second set of sulphonamides.

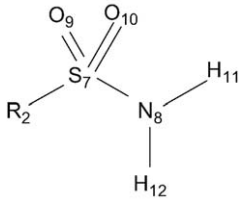

Fig. 17. Numbered common skeleton for the common SO_2NH_2 fragment.

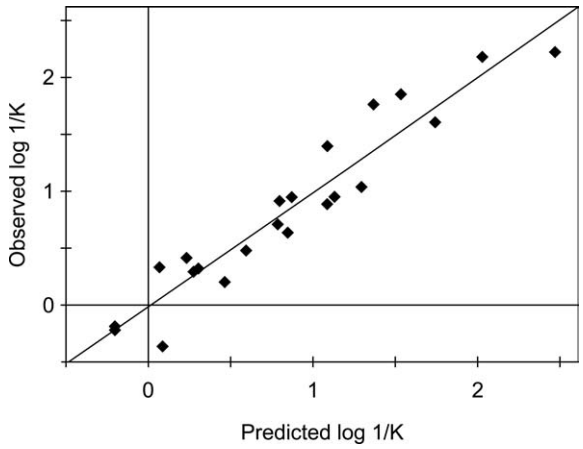

Fig. 18. Observed versus predicted log(1/K) values for the set of sulphonamides.

shows the ability of QTMS to model the electronic part of QSARs in conjunction with other descriptors.

Due to the small number of variables selected by the GA the VIP plot for the BCP properties is not given, although it should be noted that according to the authors the electronic character of the SO_2NH_2 group is crucial in determining how well binding occurs with the enzyme. Using only descriptors obtained from BCP properties on this small fragment we are able to reproduce the activities with a high degree of accuracy.

3.7. Toxicity of chlorophenols

A set of multi-substituted chlorophenol molecules have been investigated by Argese et al. [79], using six molecular descriptors chosen to represent lipophilic, electronic and steric effects: the octanol-water partition coefficient, the Hammett constant, the acid dissociation constant, the first-order valence connectivity index ($^1\chi^v$), the perimeter of the efficacious section (ΣD, a steric parameter calculated from the projection of the molecule on the plane of the aromatic ring by summing the distances between the substituents), and the melting point. In this case the Hammett constants are represented by Σσ, which are obtained by addition of the individual σ values (the *ortho* positions are assumed to have the same value as the *para* positions). The toxicity of the compounds was determined by an in vitro assay using mammalian submitochondrial particles [80]. Fig. 19 gives the common skeleton of the phenol molecules, whilst the relevant structure and activity data are given in Table S7. Here the activity, EC_{50}, is the concentration of compound required to inhibit the rate of reverse electron transfer in the assay by 50%. Table 2 gives the results of the regression, and an observed versus predicted plot for the toxicities is given in Fig. 20.

Argese et al. performed separate linear regression analyses for each of the descriptors they employed, and found the most significant descriptors to be log P and Σσ. Using the sum of the Hammett constants alone yielded an r^2 of 0.829, and a q^2 of 0.765. As a result the importance of the electronic state of the chlorophenols to the toxicity is apparent, as well as the ability

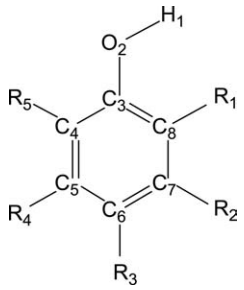

Fig. 19. Numbered common skeleton of the set of chlorophenols.

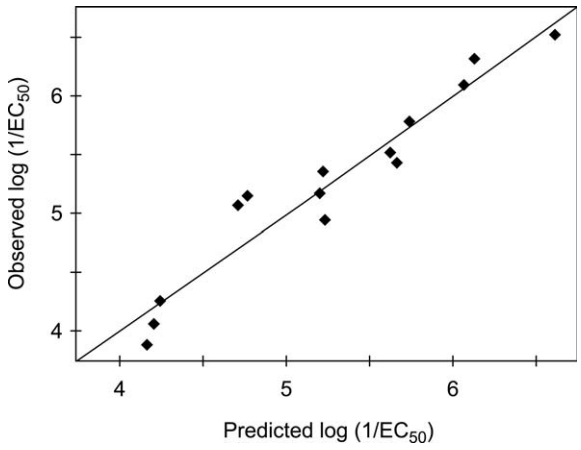

Fig. 20. Observed versus predicted log(1/EC_{50}) values for the set of chlorophenols.

of QTMS to model such effects, producing significantly higher regression statistics than the conventional electronic parameters. The most important descriptor, however, was found to be log P, which alone produced an r^2 value of 0.884, and a q^2 of 0.875. These statistics are much higher than for any of the electronic variables, and highlights the importance of hydrophobic interactions in determining the toxicity of these molecules. Argese and co-workers postulate that this probably mimics the ability of the chlorophenols to partition into the lipid bilayer of the mitochondrial membrane. Inclusion of log P into the QTMS analysis as an extra descriptor (in conjunction with the GA selected BCP properties) yields a model with $r^2 = 0.958$, and $q^2 = 0.905$. The statistics for this model are better than any obtained using the other descriptors, and once again show that the current descriptors can be used to complement lipophilic parameters for datasets where hydrophobicity and electronic effects are important. None of the steric parameters calculated exhibit good correlations with the activity, and the authors conclude that steric effects have a secondary role in determining the toxicity. Fig. 21 shows the VIP plot for the model obtained using the GA selected BCP properties with log P, and demonstrates that the hydrophobic parameter is indeed the most successful descriptor.

4. Overall discussion and conclusion

Current computers can routinely generate geometry-optimised ab initio wave functions of molecules of medicinal interest. The chemical bonds in each congeneric set of molecules are compactly described by properties defined by quantum chemical topology (QCT). These properties are called bond critical point (BCP) properties and their computation time is marginal compared to that of the wave functions. A subsequent chemometric analysis combining partial least squares with a genetic algorithm provides excellent and validated models for a range of activities of medicinal interest. The method presented is called quantum topological molecular similarity (QTMS). This method can suggest regions of the molecule responsible for a given activity. Ideally, this active region is well localised, but it can turn out to be rather diffuse or "contaminated" with bonds one would not associate with the activity.

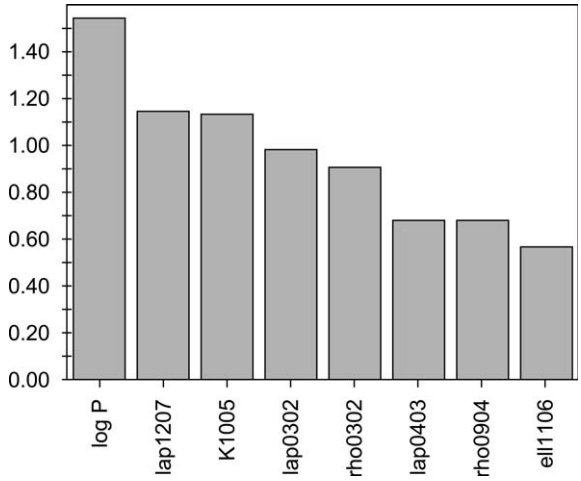

Fig. 21. VIP plot for the variables selected in the PLS analysis of the set of chlorophenols.

Although it continues to be developed and tested, QTMS has so far delivered more than two dozen of successful QSAR models for a wide variety of activities, culminating in the present study. From this work it emerged that QTMS captures the electronic effects but not the hydrophobicity nor the steric effects. However, the QTMS descriptors can of course be combined by the latter two types of descriptors, when obtained from other sources. In spite of this limitation, QTMS is an economic alternative to quantum similarity methods based on the Carbó index [81–83]. In the current method there is no need to superimpose full electron densities; it suffices to express each molecule as a QCT "quantum fingerprint" and apply a supervised algorithm to create a model. This raises the question to what extent QTMS is a 3D QSAR method. QTMS does sample electronic information from a specific 3D geometry but does not involve a superposition. The optimised geometries are local minima on the potential energy surface of the molecule. There is no guarantee or need for this minimum to be a global one. Neither have we systematically explored what the influence of conformational change is on BCP properties. However, preliminary work indicates that conformational changes in the molecule, given a fixed substituent, have a much smaller effect on BCP properties than a change in substituent. Also, in the case of the phenylbutenones a conformational study at a higher level of theory [84] showed that the active site remains unal-

tered to what was established in our original study [33]. Certainly, future work can systematically explore the issue of conformational flexibility, for example, in applications where substituents are not varied or where they induce only very small electronic changes (i.e. by being very similar or remote to the active site). In that case, each conformation could be regarded as a separate molecule, and subtle changes in the BCP properties (upon conformational change) could be observed and related to a measured activity.

The case studies presented here all assume that bonds between molecules of the same set can be matched on a one-to-one basis. This is a weak requirement given the ease with which common molecular skeletons can be identified. It is possible, in principle, to relax the common skeleton requirement totally and compare the QCT descriptors of very different molecular fragments. This more general approach is close to the idea of molecular similarity, hence the name QTMS. We point out that QCT also defines atomic properties, such as atomic volumes, energies, populations, dipole moments and higher multipole moments. Work is in progress (in a separate project) to use these atomic properties as descriptors to pinpoint bioisosterism. In that context (work to be published later) there is no need for a common molecular skeleton.

In this work we have not discussed the outcome with other possible levels of theory but this question has been addressed in other publications from this lab. The semi-empirical level AM1 does not generate (meaningful) BCPs, if at all, due to the absence of (atomic) cores in the molecular electron densities. As a result QCT descriptors cannot be drawn from this type of calculation, unless the cores are artificially added in. One could proceed with bond lengths only, as we have done for many other QSARs, but sometimes poor results are obtained because of inadequate parameterisation. A basic tenet of this article is that non-parameterised descriptors are linked and contrasted with activities of medicinal interest. This is why we restricted the reported analysis to (non-empirical) ab initio calculations (with the current basis sets). Using only a modest basis set (HF/6-31G* and HF/3-21G*) the BCP descriptors have been shown to compete with, and in many cases outperform, more conventional descriptors and 3D QSAR methods.

The objection that quantum chemists fall in the trap of "expensive ways of counting carbons" is sometimes expressed. For example, log P values of compounds such as alcohols, ketones and amides, each seen as a series with growing aliphatic side chains, can be simply related to the number of methylene groups. Obviously, there would be no need to compute accurate electron densities (or wave functions) for such series of molecules. In the sets presented here, no such trivial relationship exists, and hence the use of more sophisticated methods is warranted. The electrotopological method is such a method and one has to admit that it is much faster than quantum chemical calculations. However, the latter are real and fruitful attempts at solving the Schrödinger equation and are therefore in a position to link medicinal chemistry to an underlying physical reality. The trade-off between parsimony and veracity will remain a matter of debate. Nevertheless, one cannot deny that present day computing is so overwhelmingly abundant that avoiding ubiquitous PC Linux clusters being partially idle is almost a challenge. One may as well invest this computing time in closing the gap.

Uncited reference

[32]

Acknowledgments

The work was funded by EPSRC and (formerly) Smith-Kline-Beecham.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.ejmech.2006.03.004.

References

[1] M. Karelson, V.S. Lobanov, A.R. Katritzky, Chem. Rev. 96 (1996) 1027–1043.
[2] M. Karelson, Molecular Descriptors in QSAR/QSPR, Wiley-Interscience, New York, USA, 2000.
[3] C.M. Breneman, M. Rhem, J. Comp. Chem. 18 (1997) 182–197.
[4] C.M. Breneman, T.R. Thompson, M. Rhem, M. Dung, Comp. Chem. 19 (1995) 161–179.
[5] M. Song, C.M. Breneman, J. Bi, N. Sukumar, K.P. Bennett, S. Cramer, N. Tugcu, J. Chem. Inf. Comp. Sc. 42 (2002) 1347–1357.
[6] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Proc. Natl. Acad. Sci. USA 102 (2005) 11710–11715.
[7] K. Tuppurainen, S. Lotjonen, Mutat. Res. 287 (1993) 235–241.
[8] Y. Takahata, M.C.A. Costa, A.C. Gaudio, J. Chem. Inf. Comput. Sci. 43 (2003) 540–544.
[9] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, J.P. Tollenaere, J. Phys. Chem. A 106 (2002) 7887–7894.
[10] M.J. Sorich, R.A. McKinnon, J.O. Miners, D.A. Winkler, P.A. Smith, J. Med. Chem. 47 (2004) 5311–5317.
[11] K.C. Gross, P.G. Seybold, J. Org. Chem. 66 (2001) 6919–6925.
[12] P. Stenberg, U. Norinder, K. Luthman, P. Artursson, J. Med. Chem. 44 (2001) 1927–1937.
[13] B. Wailzer, J. Klocker, G. Buchbauer, G. Ecker, P. Wolschann, J. Med. Chem. 44 (2001) 2805–2813.
[14] T.A. Berkhout, F.E. Blaney, A.M. Bridges, D.G. Cooper, A.T. Forbes, A.D. Gribble, P.H.E. Groot, A. Hardy, R.J. Ife, R. Kaur, K.E. Moores, H. Shillito, J. Willetts, J. Witherington, J. Med. Chem. 46 (2003) 4070–4086.
[15] X. Huang, T. Liu, J. Gu, X. Luo, R. Ji, Y. Cao, H. Xue, T.F. Wong, B.L. Wong, G. Pei, H. Jiang, K. Chen, J. Med. Chem. 44 (2001) 1883–1891.
[16] L. Amat, R. Carbo-Dorca, D.L. Cooper, N.L. Allan, R. Ponec, Mol. Phys. 101 (2003) 3159–3162.
[17] E. Besalu, X. Girones, L. Amat, R. Carbo-Dorca, Acc. Chem. Res. 35 (2002) 289–295.
[18] E. Besalu, R. Carbo, J. Mestres, M. Sola, Top. Curr. Chem. 173 (1995) 31–62.
[19] X. Girones, R. Carbo-Dorca, R. Ponec, J. Chem. Inf. Comp. Sc. 43 (2003) 2033–2038.
[20] P.J. Smith, P.L.A. Popelier, J. Comp. Aided Molec. Design 18 (2004) 135–143.
[21] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[22] P.L.A. Popelier, F.M. Aicken, ChemPhysChem 4 (2003) 824–829.
[23] R.F.W. Bader, Atoms in Molecules. A Quantum Theory, Oxford Univ. Press, Oxford, GB, 1990.
[24] R.F.W. Bader, in: P.V.R. Schleyer (Ed.), Chichester, Wiley, GB, 1998, pp. 64–86.
[25] R.F.W. Bader, Chem. Rev. 91 (1991) 893–928.
[26] P.L.A. Popelier, Atoms in Molecules. An Introduction, Pearson, London, Great Britain, 2000.
[27] R.F.W. Bader, Pure Appl. Chem. 60 (1988) 145–155.
[28] R.F.W. Bader, Phys. Rev. B 49 (1994) 13348–13356.
[29] R.F.W. Bader, P.L.A. Popelier, Int. J. Quant. Chem. 45 (1993) 189–207.
[30] P.L.A. Popelier, F.M. Aicken, S.E. O'Brien, in: A. Hinchliffe (Ed.), Chemical Modelling: Applications and Theory, Vol.1, Royal Society of Chemistry Specialist Periodical Report, Ch.3, 2000, pp. 143–198.
[31] P.L.A. Popelier, P.J. Smith, in: A. Hinchliffe (Ed.), Chemical Modelling: Applications and Theory, Vol. 2, Royal Society of Chemistry Specialist Periodical Report, Ch. 8, 2002, pp. 391–448.
[32] P.L.A. Popelier, J. Phys. Chem. A 103 (1999) 2883–2890.
[33] S.E. O'Brien, P.L.A. Popelier, J. Chem. Soc., Perkin Trans. 2 (2002) 478–483.
[34] P.L.A. Popelier, U.A. Chaudry, P.J. Smith, J. Comp. Aided Molec. Design 18 (2004) 709–718.
[35] U.A. Chaudry, P.L.A. Popelier, J. Phys. Chem. A 107 (2003) 4578–4582.
[36] P.L.A. Popelier, U.A. Chaudry, P.J. Smith, J. Chem. Soc. Perkin II (2002) 1231–1237.
[37] S.E. O'Brien, P.L.A. Popelier, Quantum Molecular Similarity: Use of Atoms in Molecules derived quantities as QSAR variables, ECCOMAS Proceedings, Barcelona, Spain, 2000.
[38] S.E. O'Brien, Quantum Molecular Similarity, an Atoms in Molecules Approach, PhD Thesis, Dept. of Chemistry, UMIST, Manchester, Great Britain, 2000.
[39] B.K. Alsberg, N. Marchand-Geneste, R.D. King, Chemo. Intell. Lab. Syst. 54 (2000) 75–91.
[40] B.K. Alsberg, N. Marchand-Geneste, R.D. King, Anal. Chim. Acta 446 (2001) 3–13.
[41] U.A. Chaudry, P.L.A. Popelier, J. Org. Chem. 69 (2004) 233–241.
[42] P.J. Smith, P.L.A. Popelier, Org. Biomol. Chem. 3 (2005) (xxxx).
[43] P.J. Smith, Quantum Chemical Topological Properties as Electronic Descriptors in Quantitative Structure-Activity/Property Relationships, PhD Thesis, Dept. of Chemistry, UMIST, Manchester, Great Britain, 2003.
[44] P. Ertl, Quant. Struct. Act. Relat. 16 (1997) 377–382.
[45] L.B. Kier, Molecular structure description: the electrotopological state, Academic, San Diego, California, USA, 1999.
[46] L.B. Kier, L.H. Hall, J. Chem. Inf. Comput. Sci. 35 (1995) 1039–1045.
[47] P.L.A. Popelier, in: D.J. Wales, D.J. Wales (Eds.), Structure and Bonding. Intermolecular Forces and Clusters, Springer, Heidelberg, Germany, 2005, p. 115 (1–56).
[48] P.L.A. Popelier, Chem. Phys. Lett. 228 (1994) 160–164.
[49] N.O.J. Malcolm, P.L.A. Popelier, J. Comp. Chem. 24 (2002) 437–442.
[50] R.F.W. Bader, T.S. Slee, D. Cremer, E. Kraka, J. Am. Chem. Soc. 105 (1983) 5061–5068.
[51] S.T. Howard, O. Lamarche, J. Phys. Org. Chem. 16 (2003) 133–141.
[52] R.F.W. Bader, T.T. Nguyen-Dang, Y. Tal, Rep. Prog. Phys. 44 (1981) 893–948.
[53] J.R. Cheeseman, M.T. Carroll, R.F.W. Bader, Chem. Phys. Lett. 143 (1988) 450–458.
[54] R.F.W. Bader, H.J.T. Preston, Int. J. Quant. Chem. 3 (1969) 327–347.
[55] P.L.A. Popelier, in: M. Ford, D.J. Livingstone, J. Dearden, H. Van de Waterbeemd (Eds.), EuroQSAR2002: Designing Drugs and Crop Protectants: processes, problems and solutions, Blackwell, Oxford, GB, 2003, pp. 130–134.
[56] S.E. O'Brien, P.L.A. Popelier, J. Chem. Inf. Comp. Sc. 41 (2001) 764–775.
[57] J.B. Foresman, A. Frisch, Exploring Chemistry with Electronic Structure Methods, Gaussian Inc., Pittsburgh, USA, 1996.
[58] GAUSSIAN98: Gaussian 98, Revision A.7, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, USA, 1998.
[59] MORPHY98, http://morphy.ch.umist.ac.uk/manual.htm (1998).
[60] S. Wold, M. Sjostrom, L. Eriksson, in P.V.R. Schleyer (Ed.), Encycl. Comp. Chem., Wiley, Chichester, GB, 1998, pp. 2006–21.
[61] UMETRICS, SIMCA-P 10.0.info@umetrics.com, www.umetrics.com, Umeå, Sweden, 2002.
[62] B.M. Wise, N.B. Gallagher, PLS Toolbox and Matlab toolbox, version 3. Eigenvector Research, Manson, WA, USA, 2003.
[63] L. Amat, E. Besalu, R. Carbo-Dorca, R. Ponec, J. Chem. Inf. Comput. Sci. 41 (2001) 978–991.
[64] P.G. Mezey, Mol. Phys. 96 (1999) 169–178.
[65] A.C. Good, S.J. Peterson, W.G. Richards, J. Med. Chem. 36 (1993) 2929–2937.
[66] P.B. Timmermans, P.A. Zwieten, Drug. Res. 28 (1978) 1676–1681.
[67] B. Rouot, G. Leclerc, C.-G. Wermuth, J. Med. Chem. 19 (1976) 1049–1054.
[68] K.H. Kim, Y.C. Martin, J. Med. Chem. 34 (1991) 2056–2060.
[69] B.D. Silverman, D.E. Platt, J. Med. Chem. 39 (1996) 2129–2140.
[70] L. Amat, R. Carbo-Dorca, R. Ponec, J. Med. Chem. 42 (1999) 5169–5180.
[71] A. Da Settimo, G. Primofiore, F. Da Settimo, A.M. Marini, E. Novellino, G. Greco, C. Martini, G. Giannaccini, A. Lucacchini, J. Med. Chem. 39 (1996) 5083–5091.
[72] D. Hadjipavlou-Litina, C. Hansch, Chem. Rev. 94 (1994) 1483–1505.
[73] X. Girones, L. Amat, D. Robert, R. Carbo-Dorca, J. Comp. Aided Molec. Design 14 (2000) 477–485.
[74] W. Zhang, K.F. Koehler, P. Zhang, J.M. Cook, Drug Des. Discov. 12 (1995) 193–248.
[75] I. Tamm, K. Folkers, F.L. Horsfall, J. Exp. Med. 98 (1953) 219–229.
[76] L.H. Hall, B. Mohney, L.B. Kier, J. Chem. Inf. Comp. Sci. 31 (1991) 76–82.
[77] C. Hansch, J. Schaeffer, R. Kerley, J. Biol. Chem. 247 (1972) 4703–4710.
[78] N. Kakeya, N. Yata, A. Kamada, M. Aoki, Chem. Pharm. Bull. (Tokyo) 17 (1969) 2558–2564.
[79] E. Argese, C. Bettiol, G. Giurin, P. Miana, Chemosphere 38 (1999) 2281.
[80] E. Argese, C. Bettiol, A. Ghelli, R. Todeschini, P. Miana, Environ. Toxicol. Chem. 14 (1995) 363–368.
[81] R. Carbo, L. Leyda, M. Arnau, Int. J. Quant. Chem. 17 (1980) 1185–1189.
[82] R. Carbo-Dorca, E. Besalu, X. Girones, Advances in Quantum Chemistry 38 (2000) 1–63.
[83] R. Carbo-Dorca, D. Robert, L. Amat, X. Girones, E. Besalu, Molecular Similarity in QSAR and Drug Design, Springer, Berlin, Germany, 2000.
[84] T.T. Polgar, G. Tasi, I.G. Csizmadia, J. Mol. Struct. THEOCHEM 666–667 (2003) 131–134.